Clarifying Goals and Refining Processes Would Improve the Employee Plans Risk Assessment Program

September 2003

Reference Number: 2003-10-200

September 26, 2003

MEMORANDUM FOR COMMISSIONER, TAX EXEMPT AND GOVERNMENT
ENTITIES DIVISION

FROM: Gordon C. Milbourn III
               Assistant Inspector General for Audit (Small Business and
               Corporate Programs)

SUBJECT: Final Audit Report - Clarifying Goals and Refining Processes
               Would Improve the Employee Plans Risk Assessment Program
               (Audit # 200210045)

This report presents the results of our review of the Tax Exempt and Government
Entities (TE/GE) Division's Employee Plans (EP) Compliance Risk Assessment
Program. The overall objective of this review was to determine whether processes were
implemented to enable the EP Examinations function to achieve its goals of measuring
market segment compliance and developing strategies for improving future compliance
in the selected market segments.

In summary, the EP Examinations function implemented the Risk Assessment Program
during Fiscal Year 2002 as part of the EP Examinations Program. The EP function risk
assessment categorized qualified pension plans into 20 industry segments and 11 plan
types within each industry segment, for a total of 220 distinct market segments. EP
function management implemented a framework for the Risk Assessment Program that
enabled it to gather a significant amount of information on the risk and reliability of
different market segments in a relatively short period of time.

Although TE/GE Division management established goals for the Risk Assessment
Program, the goals are not clearly defined or measurable. For example, one of the
goals is to determine compliance levels and explore various techniques to improve
overall compliance. However, "improve overall compliance" is not defined, and there is
no measure indicating what an acceptable level of compliance is for the individual
market segments.

EP function management also established priorities for selecting examinations in each
of the market segments. Although priorities have been established, EP function
management does not have a long-term plan or strategy to accomplish these priorities.

Establishing a long-term plan or strategy and developing processes to accomplish the plan should improve the accuracy and reliability of the data used in the Risk Assessment Program. The plan should include:

- Strategies that project when examinations for specific market segments will begin and end.

- The processes that will be used to determine when education and outreach activities are appropriate.

- How many resources will be needed and whether the activities improved overall compliance.

- A determination of whether reported examination changes are consistent with the objectives of the Risk Assessment Program.

- Methods to minimize inconsistencies in the scope and depth of examinations.

- A sampling methodology that results in the most efficient use of resources.

TE/GE Division management's decision to quickly implement the Risk Assessment Program may have reduced the ability to implement additional management controls to reasonably ensure the intended Program results are achieved. Therefore, we recommended the Director, EP Examinations, establish a clear overall goal and quantitative measures for the Program, assign responsibility for developing a long-term plan and guidelines for education and outreach activities, ensure the compliance level of the market segments is based on the Program's definition of noncompliance, develop guidelines for Risk Assessment examinations to improve consistency and reliability, and evaluate various statistical sampling methods.

Management's Response: The Commissioner, TE/GE Division, generally agreed with our recommendations. The Commissioner, TE/GE Division, has developed a 3-year plan that identifies the examination Risk Assessment priorities. This plan identifies specific program objectives for each fiscal year involving the number of examinations to complete and the number of staff days applied to particular examination categories. However, EP Examinations function management believes that it would be premature to establish quantitative Program goals.

The Director, EP Examinations, will be responsible for developing a long-term plan for the Risk Assessment Program, and developing guidelines for education and outreach activities. The Director will also be responsible for using the Checksheet for EP Compliance Accomplishments to identify compliance issues based on the definition of noncompliance identified in the initial Risk Assessment Report, and developing guidelines for workpaper documentation and standard audit techniques.

The EP Examinations function did not agree that One-tailed Hypothesis testing was appropriate for validating the results under the Risk Assessment Program. Further, the EP Examinations function did not agree with our potential staff day savings of 32,678 or our recommendation to assign responsibility for evaluating various statistical sampling

methodologies to determine the method that will achieve the risk assessment goals in the most cost-effective manner. Instead, the EP Examinations function plans to explore other methods to use resources more effectively and will continue working with its Research and Analysis Staff to determine other sampling methodologies that would help validate Risk Assessment results. Management's complete response to the draft report is included as Appendix V.

Office of Audit Comment: We do not agree with EP Examinations function management's assertion that it is premature to establish quantitative Program goals. While a 3-year plan is helpful to guide the Risk Assessment Program over the next few years, EP function management still needs to define the overall goal for the Program to ensure that the short-term plans will help accomplish Program goals. In addition, although the EP function does not currently have reliable information on compliance baselines for all market segments, EP function management may not have to wait until all the baselines have been validated to determine the level at which a market segment will be considered "compliant." However, we do not plan to elevate the disagreement to the Department of the Treasury for resolution.

After the response to this report was issued, the EP Examinations function provided further clarification on the actions it will take to explore other sampling methodologies. The EP Examinations function will be bringing in a professional statistician to focus on the implementation of an effective risk assessment strategy for the next 2 to 3 fiscal years. We agree that this action will address the recommendation.

Regarding the potential outcome measure of the re-allocation of 32,678 staff days, we believe that EP function management's efforts to explore other sampling methodologies should result in more cost-effective sampling plans. The One-tailed testing alternative should still be considered because it may be appropriate for some situations. EP Examinations function management acknowledged that a One-tailed testing approach could be used to essentially validate (or invalidate) the results of the initial risk assessment but has concerns that there would be no definitive measure of the level of compliance. We do not believe that the need for such a definitive measure of the market segments has been established.

Copies of this report are also being sent to the Internal Revenue Service managers who are affected by the report recommendations. Please contact me at (202) 622-6510 if you have questions or Daniel R. Devlin, Assistant Inspector General for Audit (Headquarters Operations and Exempt Organizations Programs), at (202) 622-8500.

Table of Contents

Background	The Tax Exempt and Government Entities (TE/GE) Division's Employees Plans (EP) function is responsible for ensuring plan sponsors are in compliance with statutes and regulations governing qualified employee benefit plans and other individual account plans. The core mission of the EP Examinations function is to protect plan assets and participant benefits through a fair, objective, and effective compliance program. Currently, there are more than 700,000 qualified benefit plans with approximately 111 million participants and assets totaling more than $4 trillion.

While the number of qualified plans has increased over the last 10 years, the number of EP function examinations has declined. Frequent legislative and administrative changes present challenges for EP function management to provide sufficient resources for an adequate compliance program. These changes create inefficiencies because they require the resources of experienced employees to be shifted during the year from the Examinations Program to the Determination Program, thereby increasing the burden on managers and employees. EP function management plans to reduce some of these inefficiencies by dedicating approximately 17 percent of its total Examinations resources to perform work solely in the Determination Program.

In January 2001, the Internal Revenue Service (IRS) Commissioner directed each operating division to prepare an analysis and assessment of the compliance risks for each taxpayer segment as part of the overall IRS strategic assessment. Ideally, the assessment would be quantitative, with specific numbers showing the amount of noncompliance and an assessment of trends that could increase or decrease future noncompliance. In October 2001, the TE/GE Division issued the Report on Compliance Risk Assessments, which stated that each office within the Division had developed a definition of noncompliance and completed its market segment framework. The report also indicated that the risk assessment effort is an ongoing and permanent process to determine how to best deploy resources within the TE/GE Division.

The EP Examinations function implemented the Risk Assessment Program during Fiscal Year (FY) 2002 as part of the EP Examinations Program. The EP function risk assessment categorized qualified pension plans into 20 industry segments and 11 plan types within each industry segment, for a total of 220 distinct market segments. Each segment was assigned a risk level, depending primarily on whether there was a high, medium, or low risk of an examination change to the plan. The final risk level was based on the following data:

- Approximately 10 years of examination history on the EP Return Inventory Classification System (RICS).[1]

- Examination project reports listed on the EP Bulletin Board. These project reports summarize the results of compliance projects.

- Number of plans, total assets, and total participants for each segment.

- Collective examination experience of field personnel.

In addition, the reliability of the data used to rank the risk was assigned as high, medium, or low based on the audit rate[2] of each segment. For example, if more than 15 percent of the returns for a market segment had been examined, the reliability of the examination results for that market segment was considered to be high. Conversely, if less than 10 percent of the returns had been examined, the reliability was considered to be low.

EP function management implemented a framework for the Risk Assessment Program that enabled it to gather a significant amount of information on the risk and reliability of different market segments in a relatively short period of time. Examples of the actions taken include:

---

[1] The RICS is a computing system that provides access to EP function return and filer information.
[2] Audit rate equals the total number of returns examined divided by the total number of returns in a market segment.

- Market segments were categorized by risk and reliability using information from the RICS.

- Goals and objectives of the Program were communicated to EP function employees.

- A Risk Assessment Action Plan and a Risk Assessment Team were created. The team was established to develop the current and long-term Risk Assessment Examination plan.

- Returns from selected market segments were assigned and transmitted to the Examinations field function.

The audit was conducted in accordance with *Government Auditing Standards* at the EP Planning and Programs office located in Baltimore, Maryland, between November 2002 and May 2003. Detailed information on our audit objective, scope, and methodology is presented in Appendix I. Major contributors to the report are listed in Appendix II.

Clearer Goals Are Needed to Guide the Risk Assessment Program

Although TE/GE Division management established goals for the Risk Assessment Program and developed a framework for the market segments, the goals are not clearly defined or measurable. One of the goals of the Risk Assessment Program is to determine compliance levels and explore various techniques to improve overall compliance. However, "improve overall compliance" is not defined, and there is no measure indicating what an acceptable level of compliance is for the individual market segments.

Without specific goals and measures for the Risk Assessment Program, EP function management may not be able to determine if they are on target to meet Program goals. Also, for many of the market segments, it will take years to complete enough examinations to determine reliable risk levels. If the Program is not well-focused and monitored, there is a risk that inefficiencies or ineffectiveness may occur. For example, if there are no specific interim goals to measure progress, the EP function may not be focusing audits in the most productive areas or may not realize that its approach in a market segment is not the most effective for determining compliance levels. Later in this report, we cite examples of such program risks.

EP function management established the following measures for the Risk Assessment Program that support the IRS Corporate balanced measures:

- Business results will be achieved by the EP Examinations function using limited resources more effectively by identifying segments with a higher level of noncompliance.

- Customer satisfaction will be achieved by not examining compliant sponsors and plans as frequently.

- Employee satisfaction will be achieved by assigning more productive examinations to employees, who can see the results of their work by reviewing the Risk Assessment results posted on the IRS Intranet.

These balanced measures also need to include quantitative measures, which will enable EP function management to determine if Program goals are being achieved. For example, EP function management will need to develop a working definition of statistical compliance. Statistical compliance is the acceptable level of compliance for the individual market segments, measured using a valid statistical sampling method. An example of a working definition would be "Once a market segment reaches a certain (specified) level of compliance, it will be considered statistically compliant." EP function management could use this working definition to assist in determining the resources that can be redeployed to examine another segment. In addition, although EP function management plans to use education and outreach to assist in improving compliance, there are no Program goals or strategies for this in the Risk Assessment Program.

One reason this may have occurred is TE/GE Division management's decision to implement the Risk Assessment Program shortly after the Report on Compliance Risk Assessments was issued. The decision to quickly implement the Risk Assessment Program may have reduced the ability to implement additional management controls to reasonably ensure the intended Program results are achieved. The report was issued in October 2001 and

returns were selected using the risk assessment methodology starting in February 2002. However, the initial meeting of the EP function Risk Assessment Team did not occur until June 2002. EP function management advised us that another reason for implementing the methodology in February 2002 was because of the urgency to provide Revenue Agents with productive casework.

EP function management also advised us that quantitative measures were not established because the data used in the initial risk assessment contained inconsistencies and needed to be validated to ensure the compliance levels were accurate. As a result, EP function management thought it would be premature to establish quantitative goals with respect to improvements that could be expected.

The Office of Management and Budget Circular A-123 established, in part, that management controls are the organization, policies, and procedures used to reasonably ensure that programs achieve their intended results and that reliable and timely information is obtained, maintained, reported, and used for decision making. It also indicated that as agencies develop and execute strategies for implementing or reengineering programs and operations, they should design management structures that help ensure accountability for results.

At this early stage of Program implementation, it is critical for EP function management to establish a working definition of statistical compliance for the market segments. This compliance indicator could also serve as an overall goal. EP function management should also work toward establishing specific measures that would enable them to track progress in achieving the overall goal.

Recommendation

1. The Director, EP function, should establish a clear overall goal along with specific quantitative Program measures for the Risk Assessment Program and a method to periodically track the results.

Management's Response: A 3-year plan that identifies examination Risk Assessment priorities has been developed.

Also, specific program objectives were identified for each fiscal year involving the number of examinations to complete and the number of staff days applied to particular examination categories.

EP Examinations function management believes that it would be premature to establish quantitative Program goals for specific balanced measures improvements until they can confirm noncompliance data and consider it reliable.

Office of Audit Comment: We do not agree with EP Examinations function management's assertion that it is premature to establish quantitative Program goals. While a 3-year plan is helpful to guide the Risk Assessment Program over the next few years, EP function management still needs to define the overall goal for the Program to ensure that the short-term plans will help accomplish Program goals. In addition, although the EP function does not currently have reliable information on compliance baselines for all market segments, EP function management may not have to wait until all the baselines have been validated to determine the level at which a market segment will be considered "compliant" (statistical compliance).

Establishing a Comprehensive Long-Term Plan and Refining Processes May Improve the Risk Assessment Results

In addition to the framework established for this Program, EP function management also established the following 5 priorities (in priority order) for selecting examinations in each of the 220 market segments:

1. All Low-Reliability market segments will be examined using statistically valid samples of High-Risk cases first, followed by Medium-Risk and then Low-Risk.

2. Some large plans will be examined, based on the number of participants and/or size of assets, to determine compliance levels because of the impact that these plans have. The examinations of these plans are not planned as statistically valid samples.

3. The population of the FY 2002 market segments for which examinations have been initiated will be completed. When the lesser of 15 percent of the market segment or 400 examinations have been

completed, the assessment of risk and reliability ratings will be updated.

4. Cases will be examined (non-statistically valid samples) where significant examinations have been conducted in the past. The purpose would be to first validate the risk rating and the reliability assigned to these segments and then to identify reasons for noncompliance so that methods for addressing the noncompliance can be developed. This process will begin by conducting examinations of cases rated High-Risk/High-Reliability. Examinations will then be conducted in the following order: High-Risk/ Medium-Reliability, Medium-Risk/High-Reliability, Medium-Risk/Medium-Reliability, Low-Risk/High-Reliability, Low-Risk/Medium-Reliability.

5. The current market segments will be refined into smaller groups, as applicable, based on an analysis of the examinations done within each market segment. This "fine tuning" of the market segment will allow examinations to focus on specific issues and assist in making the segments more manageable. This process will be ongoing as data from examinations are received.

Although priorities have been established, EP function management does not have a long-term plan or strategy to accomplish these priorities. Establishing a long-term plan or strategy and developing processes to accomplish the plan should improve the accuracy and reliability of the data used in the Risk Assessment Program. EP function management should develop a strategy that addresses the following issues as it continues to implement the Program:

- When the examinations and techniques to improve compliance of the numerous market segments are projected to begin and be completed.

- What processes will be used to determine when education and outreach activities are appropriate, how many resources are needed, and whether the activities improved overall compliance.

- Whether noncompliance reported in the Risk Assessment Program is consistent with the definition of noncompliance in the Examinations Program.

- How to minimize inconsistencies in the scope and depth of the examinations so that the results of the Risk Assessment Program will be consistent and reliable.

- What sampling methodology would use resources most efficiently.

As discussed in the previous section, a long-term plan or strategy may not have been developed because of TE/GE Division management's decision to quickly implement the Risk Assessment Program.

A comprehensive long-term plan has not been developed

A key objective of the Risk Assessment Program is to design a comprehensive, long-term plan incorporating the Risk Assessment data into the planning process for all future Examinations work plans. While the EP function has incorporated the use of risk assessment data into its process of selecting returns for examination, it has not established a long-term plan for ensuring that all market segments will be incorporated into the Examinations Program. Such a plan should include deliverables and milestones for each market segment and should focus future Examinations resources into areas with the highest potential for noncompliance.

EP function management has acknowledged that validating the data in all market segments will require a significant amount of time. For FY 2003, the Risk Assessment Program was allocated 18,304 staff days or 44 percent of the total direct Examinations resources. EP function management has experienced difficulties in planning because of its need to shift resources between the Determination and Examinations Programs. A long-term plan for the Risk Assessment Program is needed to ensure these resources are focused on the highest priority areas each year and that, over time, all market segments will be effectively addressed.

EP function management has not developed processes for Risk Assessment Program education and outreach activities

The Risk Assessment Program uses a Checksheet for EP Compliance Activities (CECA) to capture the results of risk assessment examinations. The CECA includes fields for the examiner to describe issues resulting in an examination change, the cause of the error, and how the issue was resolved. The CECAs are periodically reviewed by EP function analysts to identify trends of noncompliance. A quarterly report is prepared, summarizing the noncompliance trends that are identified.

While processes have been developed for capturing the results of the examinations and analyzing noncompliance trends, processes for determining the resources necessary for education and outreach activities and the type of activities that might be the most effective in improving these noncompliance trends have not been developed. In addition, there are no processes for monitoring whether the activities actually improved compliance. Without these processes, EP function management may not know whether resources are being used effectively or if compliance is improving. EP function management advised us that resources for education and outreach efforts will be applied on an "as needed" basis and that a process had not been developed yet because the Program was in the early stages. Representatives from the Customer Education and Outreach function actively participated as members of the Risk Assessment Team, which developed the framework for the Risk Assessment Program. However, based on the trends summarized in the quarterly reports, EP function management should develop processes to determine which education and outreach activities will be a priority for the following year and allocate appropriate resources for this activity.

Noncompliance is not always consistently reported between the Examinations and Risk Assessment Programs

The TE/GE Division Report on Compliance Risk Assessments established that noncompliance exists under any of the following situations:

- Failure of the plan document to comply with requirements for tax-qualified status.

- Significant failure to operate the plan in a manner consistent with requirements for tax-qualified status.

- Failure to report and pay applicable excise taxes.

- Failure to file information returns required by the Internal Revenue Code.

- Failure to fund the plan appropriately.

However, the EP Examinations function considers all returns that require a change to be noncompliant. While the two definitions appear similar, they can be used differently in the two Programs. For example, the EP Examinations function sometimes closes a return as a change (indicating noncompliance) although the plan sponsor only corrected insignificant defects in the plan. The Risk Assessment Program has a higher standard for noncompliance because it applies only to significant failures. As stated earlier, the risk of noncompliance in the initial framework was based on several years of examination results (change rate). However, if the changes were for insignificant defects, the risk ratings for some of the market segments may be overstated.

For example, 11.5 percent of the examination changes in the initial risk assessment did not involve changes to the plan returns but only to the related returns. In these instances, the EP function identified a change to a U.S. Corporation Income Tax Return (Form 1120) or a U.S. Individual Income Tax Return (Form 1040) but did not identify a change to the Employee Benefit Plan Return (Form 5500). While these are appropriately recorded as an examination change, they do not meet the definition of noncompliance

for the Risk Assessment Program. EP function management needs to refine how risk levels are determined for the various market segments to ensure the risk levels reflect the types of noncompliance outlined in the Risk Assessment Report.

Inconsistencies in the scope and depth of examinations may adversely affect the results of risk assessment examinations

Both an EP function internal review of closed risk assessment examinations and a General Accounting Office (GAO) report[3] on EP function compliance studies expressed concerns about the accuracy of the examination results. The internal review and the GAO concluded, in part, that inconsistency in the scope and depth of examinations generally resulted in the estimates of noncompliance being inaccurate.

For example, a review of a limited sample of cases by the EP function revealed that 60 percent of the cases were closed with the wrong disposal code and that 50 percent of the wrong disposal codes were examination changes that were improperly closed as no change. As stated earlier, examination change rates are a key indicator used by EP function management to establish risk levels for market segments. The GAO report on compliance studies found that uniform standards were not developed to guide examiners in conducting the examinations and using the examination information. As a result, the examinations were not uniform and some of the data were not collected consistently.

Both reviews attributed insufficient training of examiners as being partially responsible for the differences in the scope and depth of compliance study examinations. EP function management did not establish guidelines for the scope and depth of risk assessment examinations, and the normal examination procedures indicate that the scope and depth of examinations vary depending on a number of factors. This variance can be illustrated by the following example of two

---

[3] *PRIVATE PENSIONS: IRS Can Improve the Quality and Usefulness of Compliance Studies* (GAO-02-353, dated April 2002).

examinations that were both closed as a no change, representing the low and high ends of the range. In one case, the examiner worked on the case for 2 hours and, at the other extreme, the examiner worked on the case for 158 hours.

Establishing guidelines and conducting training will help examiners involved in compliance studies collect and record information more consistently and accurately. It is important to have a consistent basis for conducting examinations, while allowing for the normal variances considered in any examination. Inconsistency in the way examinations are conducted could result in an inaccurate estimate of noncompliance for a market segment.

The sampling methodology may not be the most efficient use of resources

One of the priorities EP function management established is to examine all low-reliability market segments using statistically valid samples of high-risk cases first, followed by medium-risk cases and then low-risk cases. However, EP function management has not developed formal sampling plans or established that the selected sampling method is cost-effective. After consulting with a professional statistician, we determined that the statistical sampling methodology selected by EP function management may not be the most efficient. Using the methodology of EP function management would require 11,252 examinations to validate the change rate of all low-reliability segments (for all risk levels). However, we identified a different statistical sampling methodology that would result in much fewer examinations. If a One-tailed Hypothesis Test[4] were used, EP function management would be required to complete only 3,581 examinations to validate the risk assessment for all low-reliability segments.

---

[4] Statistical hypothesis – an opinion about a population parameter. One-tailed test – determines whether a hypothesis should be rejected. Specifically, the EP function tests would be designed to determine whether there is sufficient sample evidence to support the hypothesis that the actual examination change rates are less than the rates used for the initial risk assessment.

By using 4.26 staff days as the average length of an EP function examination, there would be a savings of 32,678 examination staff days that could be applied to other market segments or other Examinations programs.

As an illustration, the EP function selected 4 profit sharing segments during FY 2002 and has completed 909 of the planned 1,320 examinations needed for the sampling method selected. We determined that One-tailed hypothesis tests would require a sample size of only 751 examinations for the same 4 segments. Further, applying the results[5] from the 909 profit sharing examinations that have already been completed, we determined with a 95 percent confidence level that the actual examination change rate for the 4 profit sharing segments is less than the examination change rate contained in the Risk Assessment Report.

The Risk Assessment Program sampling methodology does not ensure that resources are used in the most efficient manner. Researching other sampling methodologies provides EP function management with the opportunity to focus resources on other priority areas.

Recommendations

We recommend the Director, EP Examinations:

2. Assign responsibility for developing a long-term plan for the Risk Assessment Program that includes deliverables and milestones.

Management's Response: The Director, EP Examinations, will be responsible for developing a long-term plan for the Risk Assessment Program that includes deliverables and milestones. Various milestones have already been developed and are included in the FY 2004-2005 Strategy and Program Plan. The long-term plan will be developed further after approximately 100 employees are realigned to the Determination Program.

3. Assign responsibility for developing guidelines for education and outreach activities, determining the

---

[5] We did not verify the accuracy of the examination results.

resources necessary, and developing measures to enable the EP function to determine if the education and outreach activities improve voluntary compliance.

Management's Response: The Director, EP Examinations, will develop guidelines for education and outreach activities and other relevant alternative treatments as the compliance level of each market segment is validated. A representative from Corporate Education & Outreach (CE&O) actively participates in the Risk Assessment Program and will be aware of compliance trends and issues that need to be addressed. EP Examinations function management will continue to work with the CE&O function on the development of measures to determine if education and outreach improve overall compliance.

4. Ensure that the compliance level of the market segments for the Risk Assessment Program is determined based on the Program's definition of noncompliance.

Management's Response: Only the results that are directly related to the examination of the plan under the Risk Assessment Program will be captured. The results of the risk assessment cases will be monitored to determine whether changes to the plan and its activities or reporting have occurred. The EP Examinations function has begun using the CECA to identify compliance issues based on the definition of noncompliance identified in the initial Risk Assessment Report. In addition, the EP Examinations function will begin to analyze the CECAs that capture more detailed data on the degree of changes to plan participants and plan assets.

5. Develop guidelines and training for Risk Assessment examinations to improve the consistency and reliability of the results.

Management's Response: A comprehensive review of the Examinations Program, including an analysis of examination techniques and workpaper documentation, was undertaken. Guidelines for workpaper documentation and standard examination techniques are being created. Also, EP Examinations function management is implementing a post review of closed examination cases to provide detailed

feedback to examination agents and managers regarding their cases.

6. Assign responsibility for evaluating various statistical sampling methods to determine the method that will achieve the risk assessment goals in the most cost-effective manner.

Management's Response: EP Examinations function management concluded that the One-tailed test is inappropriate and does not agree with the related outcome measure regarding the re-application of 32,678 staff days. EP Examinations function management believes that a One-tailed testing methodology would not measure the risk of an individual market segment and that a Two-tailed testing methodology is necessary to know what the risk is for a market segment within a margin of error. However, the EP Examinations function is exploring other methods to use resources more effectively and plans to work with its Research and Analysis Staff to determine other sampling methodologies that would help validate the Risk Assessment results.

Office of Audit Comment: After the response was issued, the EP Examinations function provided further clarification on the actions it will take to explore other sampling methodologies that would help validate risk assessment results. On October 1, 2003, the EP Examinations function will be bringing in a professional statistician to focus on the implementation of an effective risk assessment strategy for the next 2 to 3 fiscal years. This implementation will include the validation of risk assessment data, including populating through the examination process, appropriately sized samples in significant market segments. We agree that this action will address the recommendation. Another factor in determining the most cost-effective sampling methodology is a formal sampling plan with a clear objective of what needs to be accomplished.

Regarding the potential outcome measure of the re-allocation of 32,678 staff days, we believe that EP function management's efforts to explore other sampling methodologies should result in more cost-effective sampling plans. The One-tailed testing alternative should still be

considered because it may be appropriate for some situations. EP Examinations function management acknowledged that a One-tailed testing approach could be used to determine whether sufficient sample evidence exists to support a hypothesis that would essentially validate (or invalidate) the results of the initial risk assessment but has concerns that there would be no definitive measure of the level of compliance. We do not believe that the need for such a definitive measure of the market segments has been established.

Detailed Objective, Scope, and Methodology

The overall objective of the review was to determine whether the Tax Exempt and Government Entities Division implemented processes for the Employee Plans (EP) Compliance Risk Assessment Program to enable the EP Examinations function to achieve its goals of measuring market segment compliance and developing strategies for improving future compliance in the selected market segments. We accomplished this objective by interviewing management representatives and reviewing available documentation to:

I.   Determine whether management had established processes and guidelines to incorporate the Risk Assessment Program into the annual examination plan, including:

    A.   Documenting and communicating the overall goals of the Compliance Risk Assessment Program to the EP Examinations function staff.

    B.   Planning and assigning responsibilities for completing specific tasks necessary to complete the risk assessment project.

    C.   Describing deliverables and establishing milestones for accomplishing the assigned tasks.

    D.   Developing processes to evaluate whether the risk assessment is consistent with the goals of the Examinations Program.

II.   Determine whether management had established processes and guidelines to ensure the data used in conducting the risk assessment, and collected from the selected examinations, are both sufficient and relevant, including:

    A.   Evaluating whether the data used to assign the risk rankings are consistent with the new EP function definition of noncompliance.

    B.   Planning and executing examination projects in selected market segments including determining the extent of data gathering necessary and determining when to conduct statistically valid sampling. We consulted with a statistician during our evaluation of the statistical sampling methods used for the Risk Assessment Program.

III.   Determine whether management had established processes and guidelines to use the results from the risk assessment to plan future application of enforcement and education resources, including:

    A.   Evaluating overall results from examinations in each selected segment.

B.   Defining decision points regarding future actions based on the results of examinations.

C.   Ensuring that recommendations which merit applying additional resources are formally communicated to and incorporated in the future year work plans for the EP Examinations function and EP Customer Education and Outreach function.

Major Contributors to This Report

Daniel R. Devlin, Assistant Inspector General for Audit (Headquarters Operations and Exempt
Organizations Programs)
Nancy Nakamura, Director
James Westcott, Audit Manager
James Avery, Senior Auditor
Theodore J. Lierl, Senior Auditor
Una K. Smith, Senior Auditor
Gregory Holdeman, Auditor
Carol Rowland, Auditor

Report Distribution List

Commissioner C
Deputy Commissioner for Services and Enforcement SE
Deputy Commissioner, Tax Exempt and Government Entities Division SE:T
Director, Employee Plans, Tax Exempt and Government Entities Division SE:T:EP
Director, Employee Plans Examinations, Tax Exempt and Government Entities Division SE:T:EP:E
Chief Counsel CC
National Taxpayer Advocate TA
Director, Office of Legislative Affairs CL:LA
Director, Office of Program Evaluation and Risk Analysis RAS:O
Office of Management Controls OS:CFO:AR:M
Audit Liaison: Director, Communications and Liaison, Tax Exempt and Government Entities
        Division SE:T:CL

Outcome Measures

This appendix presents detailed information on the measurable impact that our recommended corrective actions will have on tax administration. This benefit will be incorporated into our Semiannual Report to the Congress.

Type and Value of Outcome Measure:

- Inefficient Use of Resources – Potential; 32,678 examination staff days (see page 6).

Methodology Used to Measure the Reported Benefit:

1. Using a Two-tailed testing methodology, 11,252 examinations would be required for the Internal Revenue Service (IRS) to validate all low-reliability segments.

2. Using a One-tailed testing methodology,[1] 3,581 examinations would be required for the IRS to validate all low-reliability segments.

3. This would result in 7,671 fewer examinations to validate all low-reliability segments using a One-tailed hypothesis testing methodology (11,252 – 3,581 = 7,671 examinations).

4. In Fiscal Year 2001, it took an average of 34.1 hours of Direct Examination Time to conduct an examination of 1 Employee Plans (EP) function return. Dividing the 34.1 average hours per examination by 8 hours per day results in an average of 4.26 staff days per EP function examination.

5. We multiplied the examinations that would be saved by using One-tailed testing (7,671) by the average staff days per EP function examination (4.26) and arrived at 32,678 EP Examinations function staff days available for other uses (7,671 x 4.26 staff days = 32,678 staff days).

---

[1] For the highest sample proportion at which the null hypothesis is rejected, we used 10 percent less than proportion in the null hypothesis.
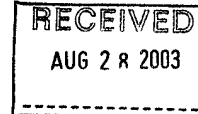
Management's Response to the Draft Report

DEPARTMENT OF THE TREASURY
INTERNAL REVENUE SERVICE
WASHINGTON, D.C. 20224

COMMISSIONER
TAX EXEMPT AND
GOVERNMENT ENTITIES
DIVISION

RECEIVED
AUG 2 8 2003

AUG 2 6 2003

MEMORANDUM FOR ASSISTANT INSPECTOR GENERAL FOR AUDIT

FROM: Evelyn A. Petschek, Commissioner
               Tax Exempt and Government Entities Division

SUBJECT: Management Response to Draft Audit Report 200210045:
               Clarifying Goals and Refining Processes Would Improve the
               Employee Plans Risk Assessment Program

I have reviewed your draft audit report on the Employee Plans (EP) risk assessment
program. In general, the report accurately describes the development methodology of
the Employee Plans Risk Assessment and offers helpful suggestions for improvement.

The initial thrust of EP Examinations through the Risk Assessment Program is to
validate the data on which the initial risk assessment was based. As noted in your
report, previous internal reviews and a General Accounting Office report have identified
past, pre-risk assessment inconsistencies in EP examination activity. While providing a
useful start for populating the Risk Assessment, we cannot use this inconsistent data as
a baseline for measuring the level of compliance in the market segments. We believe it
would be premature to establish quantitative program goals for specific balanced
measures improvements until we can confirm noncompliance data and consider it
reliable.

Further, as the report notes, the competing priorities between the EP Determinations
Program and the Examinations Program have complicated efforts to develop effective
long-term Examination Program plans. The EP organization is currently realigning and
analyzing the Determinations Program. The realignment will reflect the fact that
approximately 100 employees, who are nominally assigned to the Examinations
Program, have been consistently needed over recent years for the Determinations
Program. Thus, the realignment and other efforts to stabilize determination receipts will
diminish the need to shift EP Examination resources to the Determinations Program.
This should result in a continuity of operations that will help EP Examinations establish
and accomplish meaningful long-term program goals.

You asked that we comment on your recommended sampling methodology using "one-
tailed" testing. We do not agree that the one-tailed methodology is appropriate to
validate the results under the Risk Assessment program. Currently, we are using two-
tailed tests.

When conducting a test based on a sample, our Research and Analysis Office has indicated that we can use either a one-tailed or a two-tailed hypothesis test. One-tailed testing is normally used to prove or disprove a hypothesis. Using the EP Risk Assessment as an example, if one assumes that a segment is high risk, and high risk is defined as being greater than X% non-compliant then, as you claim, a one-tailed test is sufficient to determine whether that assumption is true (greater than X%) or false (not greater than X%). The one-tailed test would be designed to determine whether sufficient sample evidence exists to support the hypothesis that the actual examination change rates are less than the rates used for the initial risk assessment. However, the one-tailed test would not tell us what the risk is for this market segment. It still leaves EP in the position of not measuring the true compliance level within a reasonable margin of error. Rather, EP will only know that early assumptions were either true or false, but will still be left with no definitive measure of the levels of compliance.

On the other hand, the two-tailed test will allow EP to know what the risk is for this market segment (within a margin of error). By using the two-tailed test, EP will know these compliance levels with some precision.

Page 12 of the report also suggests that EP use the results from the 909 profit sharing examinations that we have already completed without waiting for the results from the rest of the cases in the sample. We have consistently found that cases that are still open as we near the end of the sample study are likely to be "change" cases. If we do not wait for these results, the sample and estimate will only represent the easier, no-change cases and miss the true estimate of non-compliance. It is an inappropriate technique, given our knowledge of case processing, for both one-tailed and two-tailed tests.

In summary, we conclude that the one-tailed test is inappropriate, and we do not agree with the related assumptions and conclusions regarding the re-application of 32,678 staff days. We are exploring other methods to use our resources more effectively, including conducting limited focus examinations. We are also working with our Research and Analysis Staff to determine other sampling methodologies that would help us validate the Risk Assessment results.

Our comments on your recommendations in the report are as follows:

Recommendation 1:

The Director, EP function, should establish a clear overall goal along with specific quantitative Program measures for the Risk Assessment Program and a method to periodically track the results.

Corrective Action:

We have developed a three-year plan that identifies our examination Risk Assessment priorities. We also identified specific program objectives for each fiscal year involving the number of examinations to complete and the number of staff days applied to particular examination categories.

Our EP Examinations personnel already track the results of these examinations and make the necessary adjustments to deliver our workplan priorities.

Implementation Date: Completed.

Responsible Official: Not applicable. Corrective action is completed.

Corrective Action Monitoring Plan: Not applicable. Corrective action is completed.

<u>Recommendation 2</u>:

Assign responsibility for developing a long-term plan for the Risk Assessment Program that includes deliverables and milestones.

Corrective Action:

The Director, EP Examinations will be responsible for developing a long-term plan for the Risk Assessment Program that includes deliverables and milestones. We have already developed various milestones that we have included in our FY 2004-2005 Strategy and Program Plan. We will further develop the long-term plan after we realign approximately 100 employees to the Determination Program. The realignment and other efforts to stabilize determination receipts should help EP Examinations establish and accomplish meaningful long-term program goals.

Implementation Date: January 31, 2004

Responsible Official: Director, EP Examinations

Corrective Action Monitoring Plan:

We will monitor this corrective item through the Treasury Department's Joint Audit Management Enterprise System.

<u>Recommendation 3:</u>

Assign responsibility for developing guidelines for education and outreach activities, determining the resources necessary, and developing measures to enable the EP

4

function to determine if the education and outreach activities improve voluntary compliance.

Corrective Action:

As we validate the compliance level of each market segment, we will develop guidelines for education and outreach activities, and other relevant alternative treatments. A representative from Corporate Education & Outreach (CE&O) actively participates in the Risk Assessment Program and will be aware first-hand of compliance trends and issues that need to be addressed. We currently allocate staffing to CE&O activities. In the future, we will direct some of those staffing resources towards addressing the results of our validated risk assessment results. We will continue to work with CE&O to develop measures to determine if education and outreach improve overall compliance.

Implementation Date: September 30, 2004

Responsible Official: Director, EP Examinations

Corrective Action Monitoring Plan:

We will monitor this corrective item through the Treasury Department's Joint Audit Management Enterprise System.

Recommendation 4:

Ensure that the compliance level of the market segments for the Risk Assessment Program is determined based on the Program's definition of noncompliance.

Corrective Action:

Under the Risk Assessment Program, we will capture only the results that are directly related to the examination of the plan.

We monitor the results of the risk assessment cases to determine whether changes to the plan and its activities or reporting have occurred. We have begun using the CECA (Checksheet for EP Compliance Accomplishments) checksheet to identify compliance issues based on the definition of noncompliance identified in the initial Risk Assessment Report. We will begin to analyze the CECA checksheets that capture more detailed data on the degree of changes to plan participants and plan assets.

Implementation Date: October 1, 2003

Responsible Official: Director, EP Examinations

5

Corrective Action Monitoring Plan:

We will monitor this corrective item through the Treasury Department's Joint Audit
Management Enterprise System.

Recommendation 5:

Develop guidelines and training for Risk Assessment examinations to improve the
consistency and reliability of the results.

Corrective Action:

We have undertaken a comprehensive review of our Examination Program, which
includes an analysis of examination techniques and workpaper documentation. We are
creating guidelines for workpaper documentation and standard audit techniques. We
are also implementing a post-review of closed examination cases to provide detailed
feedback to our examination agents and managers regarding their cases.

Implementation Date: July 31, 2004

Responsible Official: Director, EP Examinations

Corrective Action Monitoring Plan:

We will monitor this corrective item through the Treasury Department's Joint Audit
Management Enterprise System.

Recommendation 6:

Assign responsibility for evaluating various statistical sampling methods to determine
the method that will achieve the risk assessment goals in the most cost-effective
manner.

Corrective Action:

We do not agree with this recommendation for the reasons cited on pages 1 and 2 of
our response. A representative from TE/GE's Research and Analysis Staff is a member
of the Risk Assessment Team and participates in the decisions regarding the
appropriateness of the statistical sampling methods.

Implementation Date: Not applicable.

Responsible Official: Not applicable.

6

Corrective Action Monitoring Plan: Not applicable.

If you have any questions, please contact Preston Butcher at (410) 962-4092.